

Smart Memory

www.huawei.com

Bill Lynch
Sailesh Kumar
and Team

HUAWEI TECHNOLOGIES Co., Ltd.



Overview

- **High Performance Packet Processing Challenges**
- **Solution – Smart Memory**
- **Smart Memory Architecture**

Packet Processing Workload Challenges

- **Sequential memory references**
 - › For lookups (L2, L3, L4, and L7)
 - › Finite automata traversal
- **Read-modify-write**
 - › Statistics, counters, token-bucket, mutex, etc
- **Pointer and link-list management**
 - › Buffer management, packet queues, etc
- **Traditional implementations use**
 - › Commodity memory to store data
 - › NPs and ASICs to process data in memory

**Tons of memory references
and minimal compute**

Performance Barriers:

1. **Memory and chip I/O bandwidth**
2. **Memory latency**
3. **Lock for atomic access**

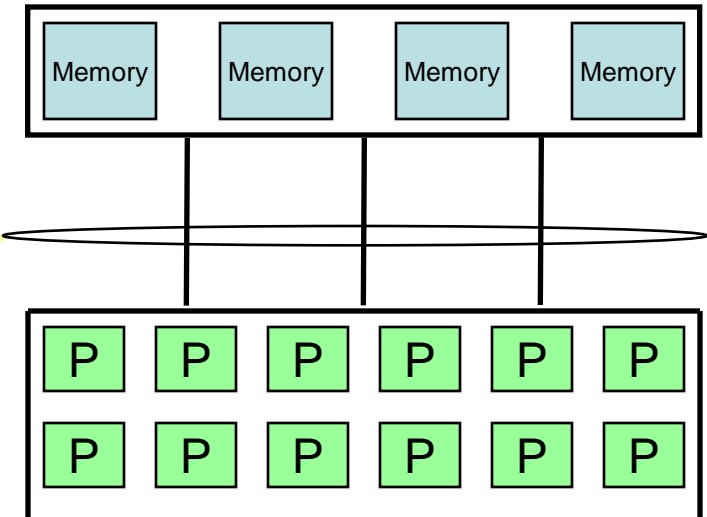


Illustration of Performance Barrier I

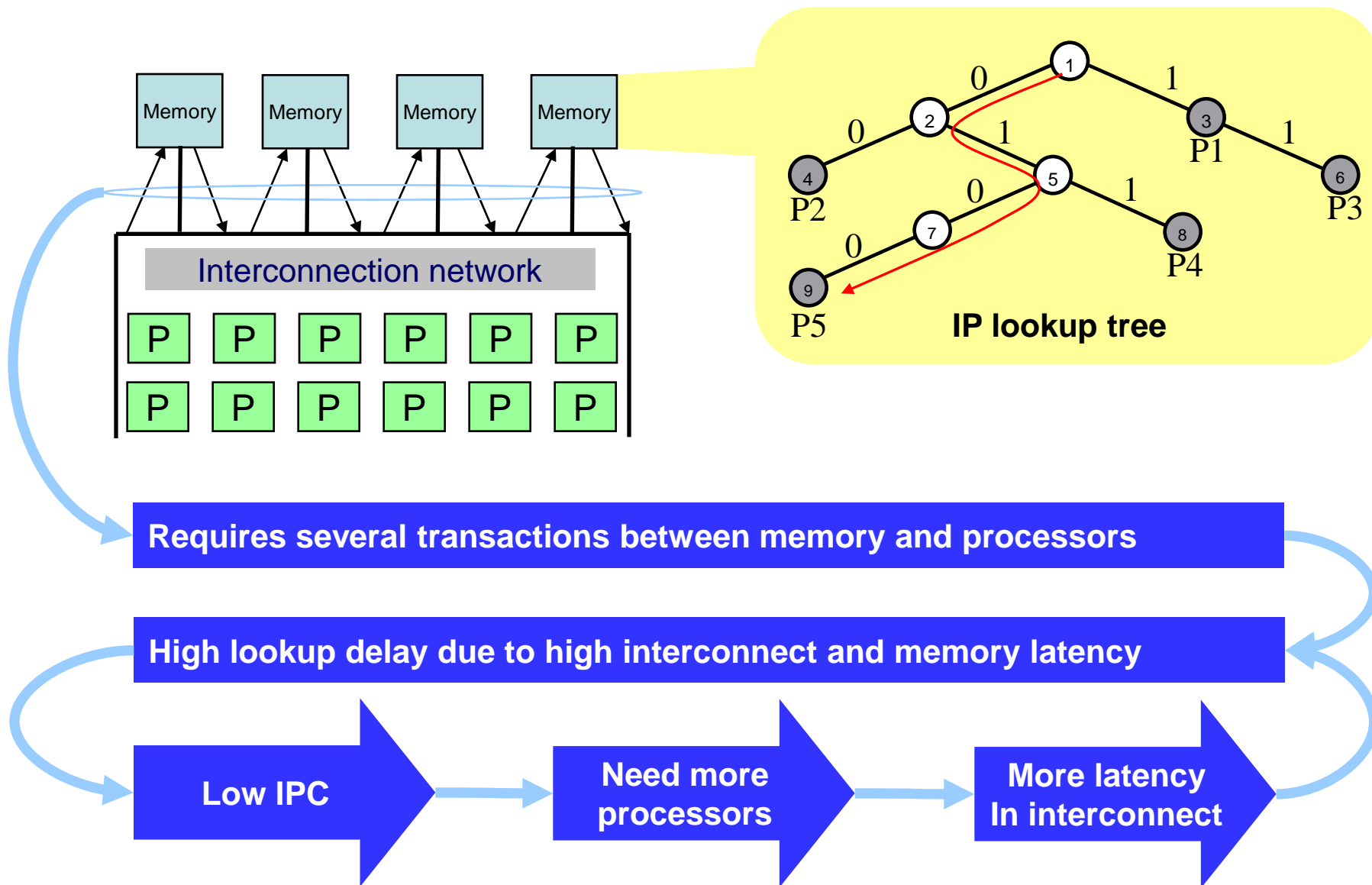
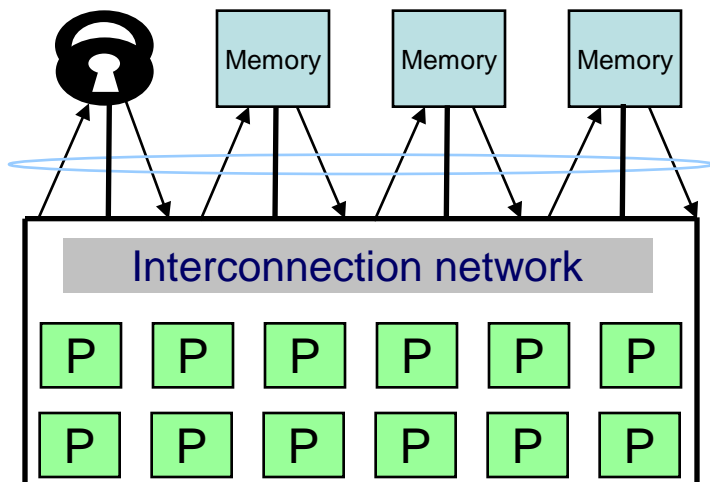
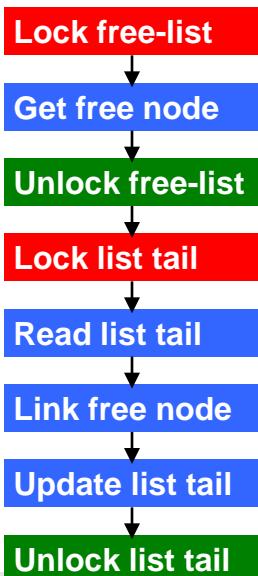


Illustration of Performance Barrier II

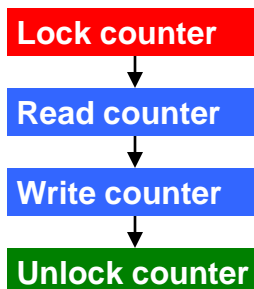


- Lookups are read-only so relatively easy
- Link-list, counters, policers, etc are **read-modify-write**
- Requires per memory address **lock** in multi-core systems

Enqueue
Dequeue



Counters



Locks often kept in memory

Requires another transaction

Adds significant latency

Single queue or single counter operations are **extremely slow**

Overview

✓ High Performance Packet Processing Challenges

- › Memory bandwidth and latency
- › Limited I/O bandwidth

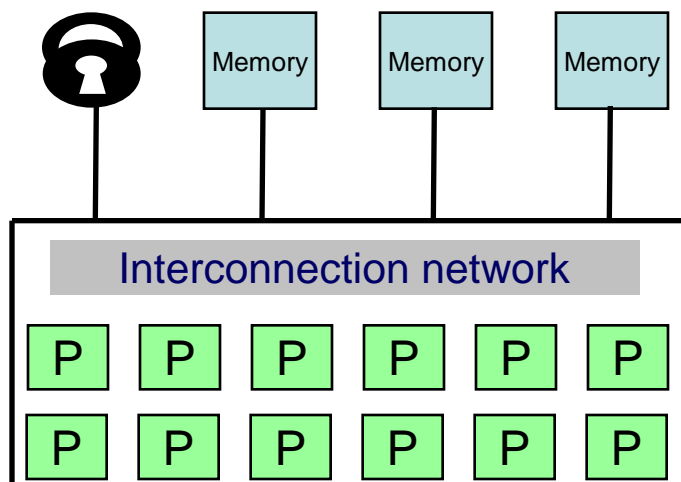
● Solution – Smart Memory

- › Attach simple compute with data
- › Attach lock with data
- › Enable local memory communication

● Smart Memory Architecture

- › Hybrid memory – eDRAM + DDR3-DRAM
- › Serial I/O

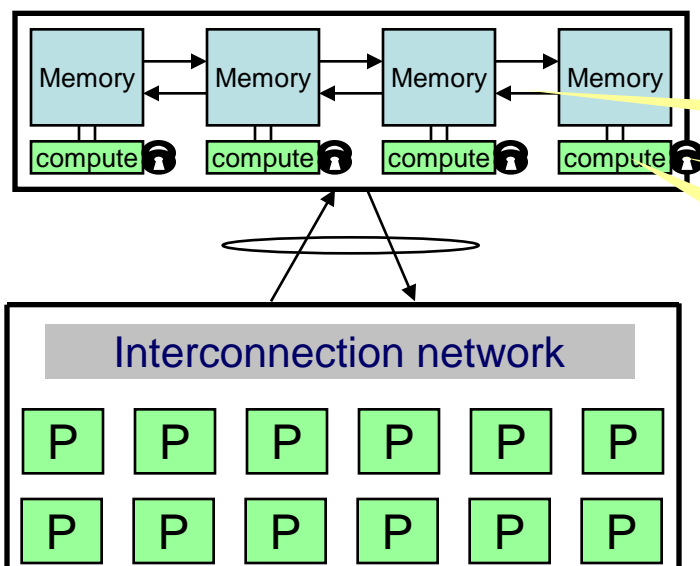
Introduction to Smart Memory



- **What is the real problem?**

- › Compute occurs far away from data
- › Lock acquire/release occurs far from data

Fortunately, compute for packet processing jobs are very modest!



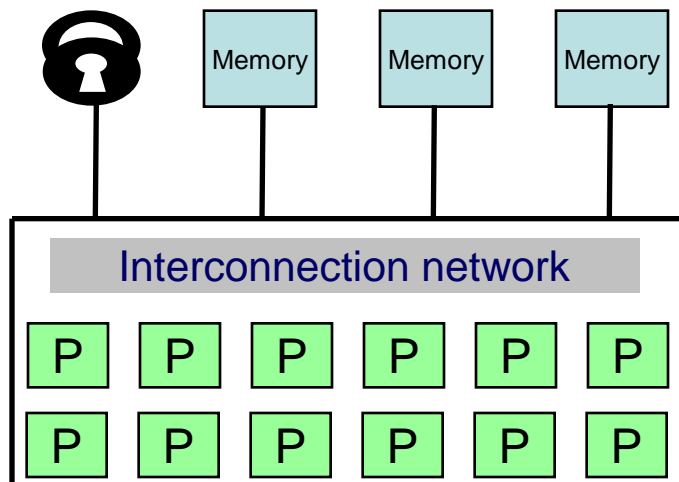
- **Solution: Make memory smarter by:**

Enabling local communication

Managing lock close to data

Keeping compute close to data

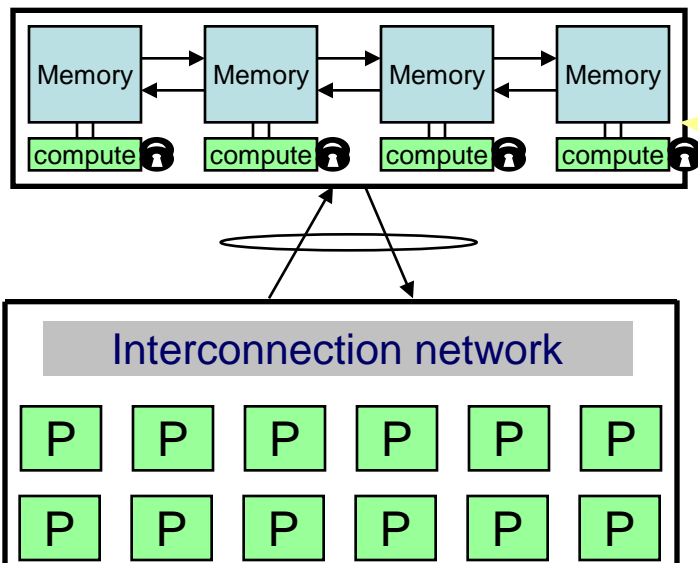
Introduction to Smart Memory



- **What is the real problem?**

- › Compute occurs far away from data
- › Lock acquire/release occurs far from data

Fortunately, compute for packet processing jobs are very modest!



Smart Memory Advantages (Get more off fewer transactions!)

1. Lower I/O bandwidth
2. Lower processing latency
3. Higher IPC
4. Significantly higher single counter/queue performance

Overview

✓ High Performance Packet Processing Challenges

- › Memory bandwidth and latency
- › Limited I/O bandwidth

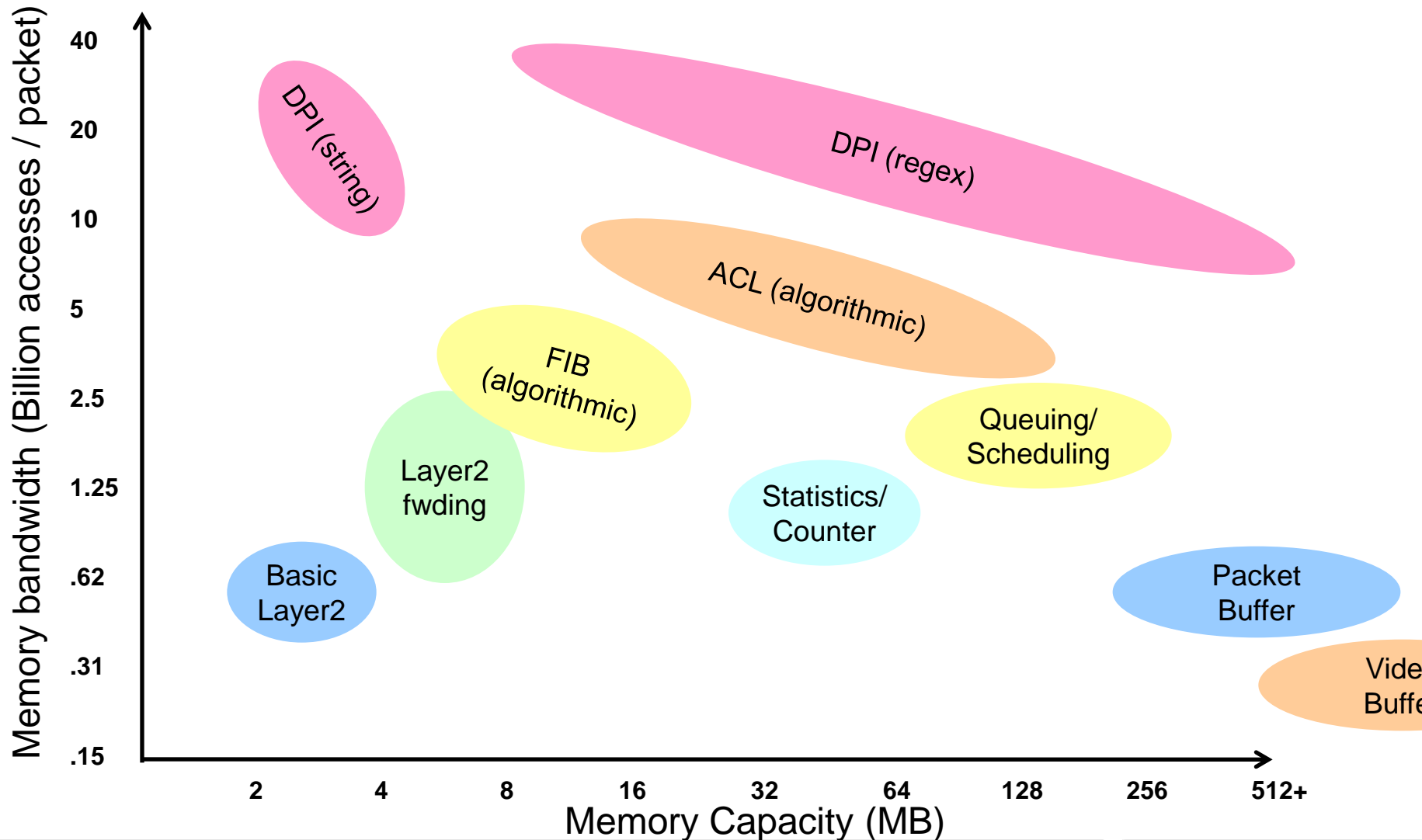
✓ Solution – Smart Memory

- › Attach simple compute with data
- › Attach lock with data
- › Enable local memory communication

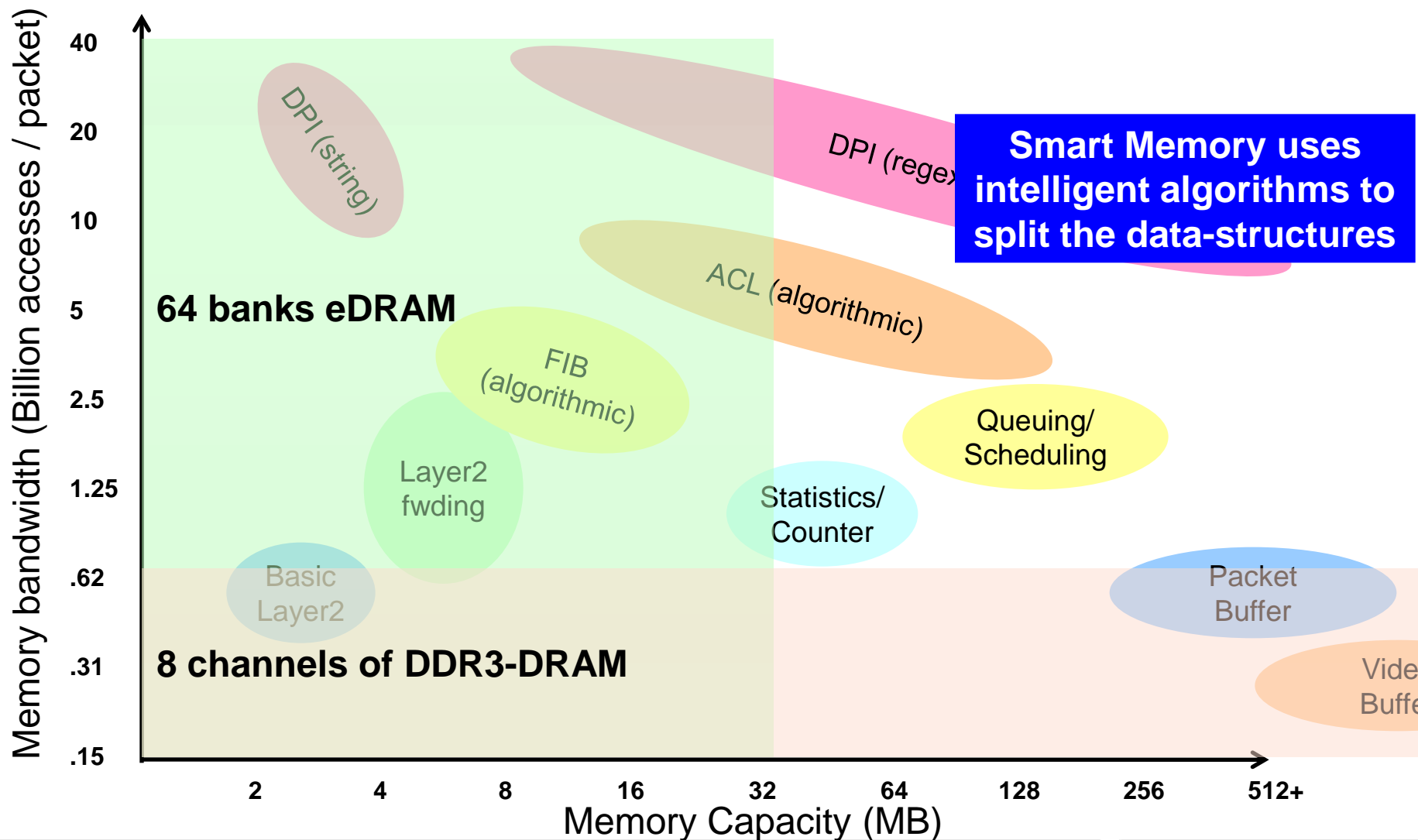
● Smart Memory Architecture

- › Hybrid memory – eDRAM + DDR3-DRAM
- › Serial chip I/O

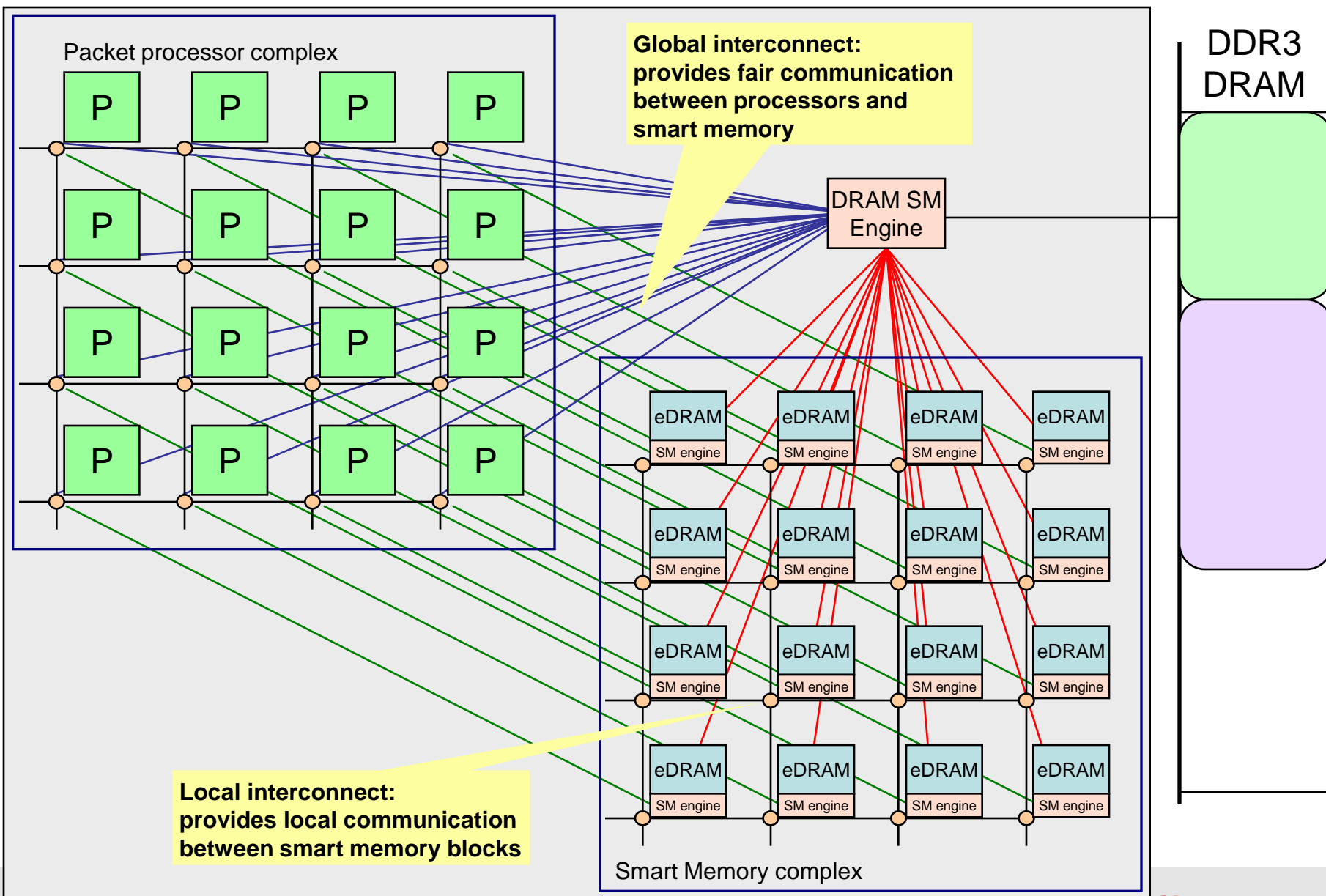
Smart Memory Capacity and Bandwidth @100G



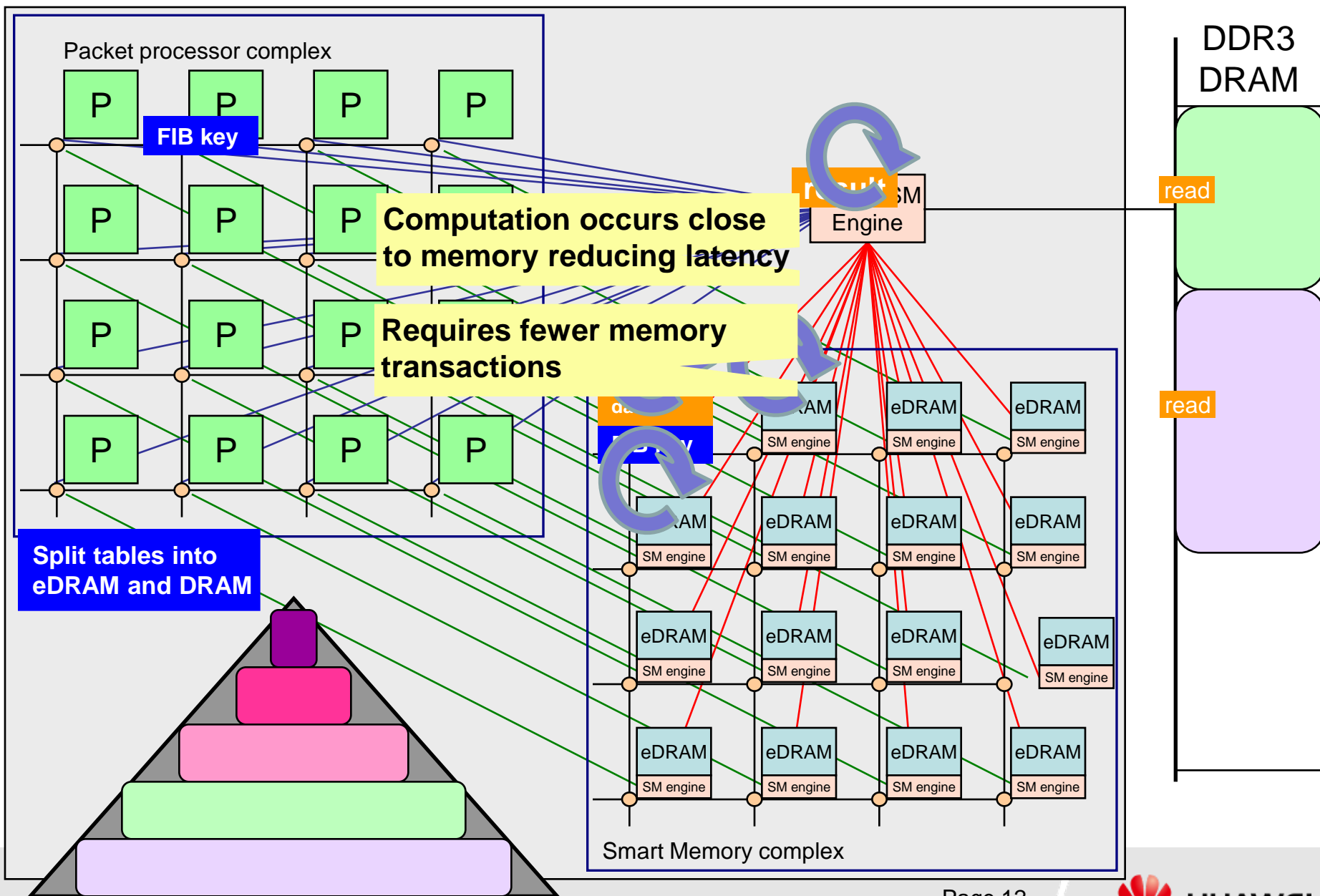
Smart Memory Capacity and Bandwidth @100G



Smart Memory High Level Architecture

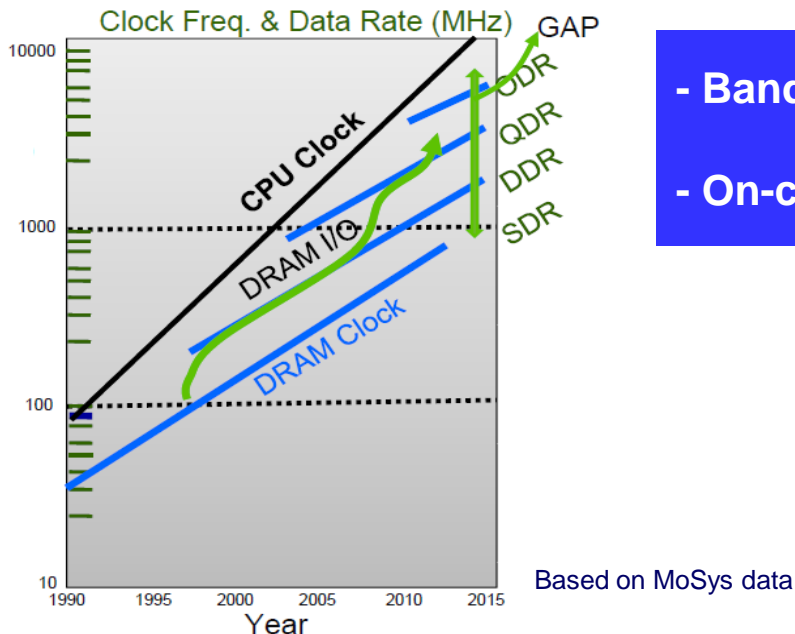


Smart Memory High Level Architecture



I/O Technology Choice in Smart Memory

- Smart Memory reduces the chip I/O bandwidth significantly
- How to further optimize it?



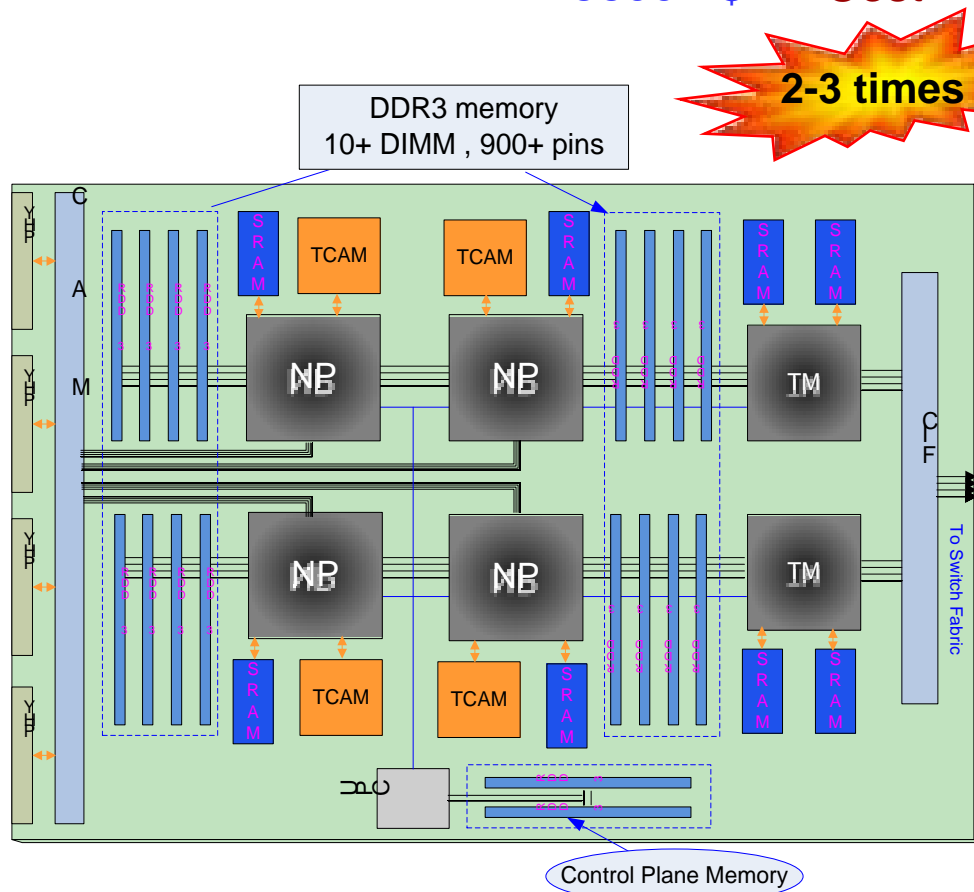
- Bandwidth, latency and I/O bandwidth gap is growing
- On-chip bandwidth is much higher than memory I/O

Smart Memory use serial I/O

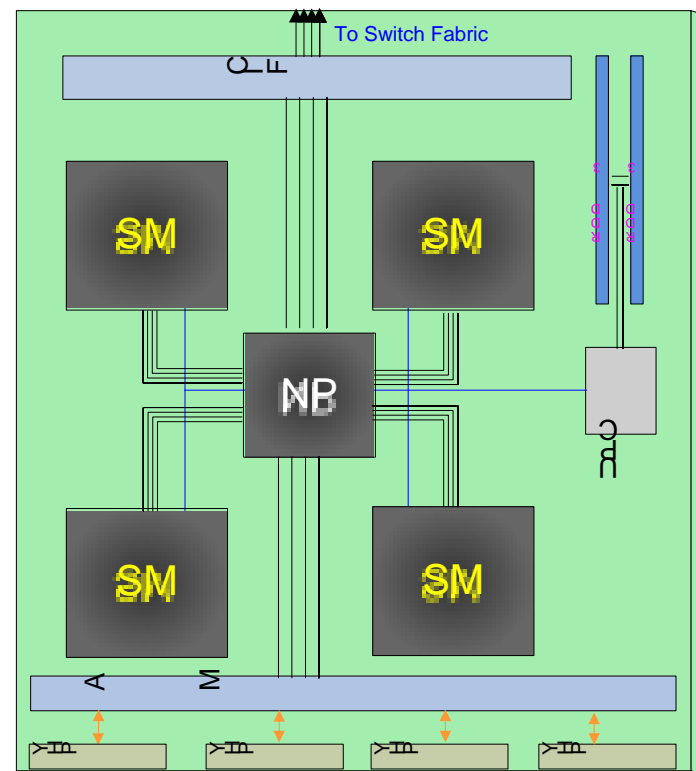
- 4X throughput than RLDRAM and QDR
- 3X fewer pins than DDR3 and DDR4
- 2.5X reduces I/O power

High Speed Line Card with Smart Memory

	540+ w	Power	212- w	
	472+ cm ²	Area	148 cm ²	
Traditional Line Card	5600+ \$	Cost	2520 \$	Line Card with SM



2-3 times



Concluding Remarks

- **Packet Processing Bottlenecks**

- › Data away from compute
- › I/O and memory bandwidth

- **Smart Memory**

- › Keep compute close to data
- › Keep locking close to data
- › Provide inter-memory connect

- **Advantages**

- › Reduced chip I/O bandwidth
- › High performance and low latency
- › Feature rich, flexible and programmable
- › Lower cost
- › One chip for several functions

Thank You

www.huawei.com